APPLICATIONS OF NEXT-GENERATION SEQUENCING

Charting histone modifications and the functional organization of mammalian genomes

Vicky W. Zhou[*‡§||¶], *Alon Goren*[*‡§¶] *and Bradley E. Bernstein*[*‡§]

Abstract | A succession of technological advances over the past decade have enabled researchers to chart maps of histone modifications and related chromatin structures with increasing accuracy, comprehensiveness and throughput. The resulting data sets highlight the interplay between chromatin and genome function, dynamic variations in chromatin structure across cellular conditions, and emerging roles for large-scale domains and higher-ordered chromatin organization. Here we review a selection of recent studies that have probed histone modifications and successive layers of chromatin structure in mammalian genomes, the patterns that have been identified and future directions for research.

*Howard Hughes Medical Institute and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, 02114, USA. ‡Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA. §Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, 02142, USA. ||Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, 02115, USA. ¶These authors contributed equally to this work. Correspondence to B.E.B. e-mail: Bernstein.Bradley@mgh.harvard.edu
doi:10.1038/nrg2905
Published online
30 November 2010

The initial sequencing of the human genome a decade ago marked a shift away from a gene-centric paradigm and prompted many new lines of genome-scale investigation. An important emerging area relates to the packaging of DNA into chromatin and, specifically, how cell type-specific chromatin organization enables differential access to and activity of regulatory elements and the manifestation of unique cellular phenotypes.

Eukaryotic chromatin structure can be viewed as a series of superimposed organizational layers[1,2] (FIG. 1). At the root are the DNA sequence and its direct chemical modification by cytosine methylation[3]. The DNA is folded into nucleosomes — the fundamental units of chromatin — that comprise approximately 147 bp of DNA wrapped around a histone octamer. The nucleosomal histones H2A, H2B, H3 and H4 can be chemically modified and exchanged with variants. The nucleosome positions along with histone variants and modifications make up the primary structure of chromatin. Finally, three-dimensional models of chromatin in nuclei are now being developed with increasing precision and propose that there are additional sophisticated layers of genome regulation through higher-order organization and nuclear compartmentalization.

With increasing knowledge of chromatin structure and its attributes at different genomic loci and in various cell types comes the challenge to elucidate which elements and regulatory processes determine this structure. Specific chromatin configurations may be dictated by DNA sequence, DNA methylation patterns, transcription factors and other regulatory proteins, and transcriptional activity, and may be maintained through epigenetic controls that are rooted in the chromatin machinery[4]. Sequence features, such as CpG islands, promoters and repetitive elements, tend to assume characteristic modification patterns and chromatin states. These patterns result from complex mechanisms involving *trans*-acting factors that are subject to intense investigation but remain poorly understood[4–6]. These distinctive chromatin configurations facilitate targeting of transcription factors and regulatory machinery to active genomic elements in mammalian genomes. As the chromatin patterns at a particular locus are intimately related to underlying regulatory processes, they may vary markedly with cellular context. In particular, chromatin is heavily influenced by transcription factor networks and transcriptional processes, which extensively harness chromatin modifiers and nucleosome remodellers[7]. In certain cases, environmental and stochastic events may invoke stable alterations in chromatin patterns, although our understanding of the output of such effects remains minimal[8].

Large-scale mapping of histone modifications and related structures has emerged as a powerful means for characterizing the determinants and the functional consequences of chromatin structure. Here, we review recent studies that have applied technologies such as chromatin immunoprecipitation followed by sequencing (ChIP–seq) to interrogate chromatin structure across the genome in diverse cell types, with an emphasis on mammalian models. We briefly present the technological

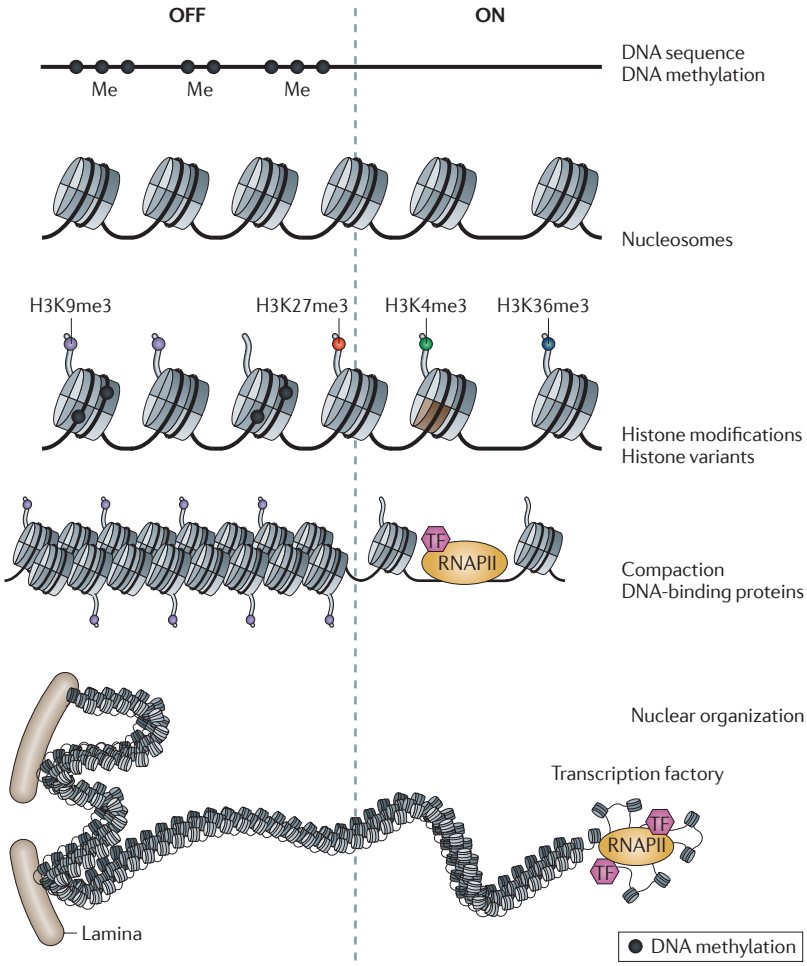

Figure 1 | **Layers of chromatin organization in the mammalian cell nucleus.** Broadly, features at different levels of chromatin organization are generally associated with inactive (off) or active (on) transcription. From the top, genomic DNA is methylated (Me) on cytosine bases in specific contexts and is packaged into nucleosomes, which vary in histone composition and histone modifications (for example, histone H3 lysine 9 trimethylation (H3K9me3)); these features constitute the primary layer of chromatin structure. Here, different histone modifications are indicated by coloured dots and histone variants such as H2A.Z are brown. DNA in chromatin may remain accessible to DNA-binding proteins such as transcription factors (TFs) and RNA polymerase II (RNAPII) or may be further compacted. Chromatin can also organize into higher-order structures such as nuclear lamina-associated domains and transcription factories. Each layer of organization reflects aspects of gene and genome regulation.

developments that have punctuated the shift from a gene-centric to genome-wide view. Then we discuss our current knowledge of primary chromatin structure, focusing on the global patterns, functions and dynamics of histone modifications that overlay sequence features such as promoters, enhancers and gene bodies. Finally, we will discuss notable recent studies that illuminate the link between histone modifications and higher-order chromatin domains.

CpG island
A genomic region enriched for CpG dinucleotides that often occurs near constitutively active promoters. Mammalian genomes are otherwise depleted of CpGs owing to the preferential deamination of methylated cytosines.

From gene-centric to genome-wide

For the past several decades, chromatin biology has been guided by a succession of methods for probing features such as chromatin accessibility; DNA methylation; the location, composition and turnover of nucleosomes; and the patterns of post-translational histone modifications. Technological advances in microarrays and next-generation sequencing have enabled many of these assays to be scaled genome-wide. Notable examples include: the DNase I–seq[9,10], FAIRE–seq[11] and Sono–seq[12] assays for chromatin accessibility; whole-genome and reduced-representation bisulphite sequencing (BS-seq)[13,14] and MeDIP-seq[15] assays for DNA methylation; and the MNase–seq[16,17] and CATCH–IT[18] assays for elucidating nucleosome position and turnover, respectively. These technologies and their integration have been extensively reviewed elsewhere[19,20]. In this section, we focus on histone modifications and, in particular, on how genome-wide ChIP–seq-mapping studies have enhanced our understanding of the chromatin landscape.

Mapping histone modifications genome-wide. Although ChIP has been used since 1988 (REF. 21) to probe chromatin structure at individual loci, its combination with microarrays and, more recently, next-generation sequencing has provided far more precise and comprehensive views of histone modification landscapes, which have highlighted roles for chromatin structures across diverse genomic features and elements that were not appreciated in targeted studies. The basis of ChIP is the immunoprecipitation step, in which an antibody is used to enrich chromatin that carries a histone modification (or other epitope) of interest. In ChIP–seq, next-generation technology is used to deep sequence the immunoprecipitated DNA molecules and thereby produce digital maps of ChIP enrichment (BOX 1). An example is the comprehensive work by Keji Zhao's group to profile 39 different histone methylation and acetylation marks genome-wide in human $CD4^+$ T cells[22,23]. These maps and similar data sets[24–26] have associated particular modifications with gene activation or repression and with various genomic features, including promoters, transcribed regions, enhancers and insulators (FIG. 2). These and subsequent studies highlight the value of comprehensive and less-biased sequencing approaches for testing the generality of insights gleaned through gene-specific studies, as well as for identifying altogether new associations and biological phenomena.

Integrating ChIP–seq maps. The expanding body of chromatin data in the public domain has fostered many computational efforts that aim to integrate different data types, identify novel relationships among histone modifications and related chromatin structures, and develop new hypotheses regarding the regulatory functions of these chromatin features. Integration of histone modification maps with chromatin accessibility, nucleosome positions, transcription factor binding, RNA expression and sequence-based genome annotations is providing increasingly unified views of chromatin structure and function[17,19,27].

Two recent studies have presented innovative approaches for integrating genome-wide chromatin maps[28,29], both of which were demonstrated on a compendium of ChIP–seq data for human $CD4^+$ T cells[22,23].

Box 1 | **ChIP–seq: current limitations and future progress**

Enabled by technological advances and plummeting costs of DNA sequencing, genome-wide maps for histone modifications and related chromatin structures are being generated at ever increasing rates. Given this expanding reliance on chromatin immunoprecipitation followed by sequencing (ChIP–seq) technology and data, there is a need for the uniform implementation of data standards. The Encyclopedia of DNA Elements (ENCODE) Project[27] and the National Institutes of Health Roadmap for Epigenomics[116] have established standards for experimental procedures, documentation and quality controls that are intended to ensure the quality and facilitate the portability, interpretation and integration of functional genomic data.

Questions still remain at the level of biological interpretation of ChIP–seq data. Inherent to ChIP technology is the fact that it reports on the relative enrichment of a modification across a population of cells. Accordingly, it cannot discern the absolute level of these modifications, that is, what fraction of histone tails at a given locus is modified, and it may be confounded by cellular heterogeneity. The magnitude of enrichment signal is also an important consideration. A few modifications typically show enrichments of 10- to 100-fold and thereby offer particularly reliable metrics. Signals for many other epitopes tend to be subtler, but could be equally biologically important. In such cases, it can be difficult to discern whether perceived differences reflect technical issues such as inefficient immunoprecipitation, or true biological phenomena. Significant trends can often be detected through composite analysis of hundreds of genes or elements, but biological conclusions should be made with care when overall differences in magnitude are incremental. Although these limitations are starting to be addressed by improved ChIP–seq procedures that increase sensitivity and reliability, there is an urgent need for orthogonal approaches.

ChIP–seq
Chromatin immunoprecipitation followed by sequencing. A method for mapping the distribution of histone modifications and chromatin-associated proteins genome wide that relies on immunoprecipitation with antibodies to modified histones or other chromatin proteins. The enriched DNA is sequenced to create genome-wide profiles.

DNase I–seq
DNase I digestion followed by sequencing. A method that distinguishes open chromatin regions based on their hypersensitivity to DNase I digestion. Sequencing these genomic fragments can generate genome-wide maps of chromatin accessibility.

FAIRE–seq
Formaldehyde Assisted Isolation of Regulatory Elements followed by sequencing exploits the solubility of open chromatin in the aqueous phase during phenol–chloroform extraction to generate genome-wide maps of soluble chromatin.

Sono–seq
Sonication followed by sequencing. A technique that relies on the increased sonication efficiency of open crosslinked chromatin to identify regions of increased accessibility genome-wide.

MNase–seq
Micrococcal nuclease digestion followed by sequencing. A method that distinguishes nucleosome positioning based on the ability of nucleosomes to protect associated DNA from digestion by micrococcal nuclease. Protected fragments are sequenced to produce genome-wide maps of nucleosome localization.

CATCH–IT
Covalent Attachment of Tags to Capture Histones and Identify Turnover is an assay for measuring nucleosome turnover kinetics genome-wide by metabolically labelling histones and profiling labelled DNA using microarrays.

Hon *et al.* applied a pattern-finding algorithm called ChromaSig to identify combinations of histone modifications at predetermined classes of regulatory loci, including promoters and enhancers. After validating that their approach identified known associations between modifications and expression levels, they applied it to regions outside these elements and subsequently identified distinct chromatin signatures associated with exons and large-scale repressed regions. Ernst *et al.* used a multivariate Hidden Markov Model to discover biologically meaningful combinations *a priori*. They discovered 51 distinct chromatin states that could be subdivided according to current genome annotations, including several promoter-associated, enhancer-associated and repressed states. This unbiased approach revealed the high information content provided by combinatorial modification patterns. It also confirmed striking functional distinctions between histone methylation marks that affect different histone residues or with different degrees of chemical modification (mono-, di- or trimethylation). By contrast, the functional correlates of histone acetylation marks seemed to be less dependent on the specific residues involved and instead depended on the overall degree of acetylation, consistent with previous studies in yeast[30,31].

Although their findings are largely consistent with prior knowledge of histone modification functions, these studies are important for their forward-looking approaches to developing algorithms that integrate increasingly vast bodies of functional genomic data into coherent biological views. A key future direction will be an equally systematic characterization of chromatin-associated proteins, including the regulators that modify and otherwise interact with histones. Such data could facilitate perturbation of specific chromatin structures to thereby yield insights into their functions. Although this goal will be technically challenging, a recent study in *Drosophila melanogaster* that mapped dozens of chromatin proteins, and thereby partitioned the genome based on their combinatorial binding patterns, provides a potential path forward[32].

Histone modifications across sequence elements

In this section, we review the types and patterns of histone modifications that have been linked to major functional genomic elements, discuss their dynamics through cell differentiation and development, and touch on functional studies that are beginning to give a mechanistic grounding to these observed patterns.

High- and low-CpG content promoters. Although mammalian promoter regions vary considerably in their positional relationships to genes, the DNA sequence proximal to the transcriptional start site (TSS) of a gene (for example, the region ± 2 kb) is frequently regarded as a proxy. The patterns of histone modification across such regions offer insights into the regulatory state of promoters and genes, and have revealed important paradigms of gene regulation.

Mammalian promoters can be classified according to their sequence content and this has proved useful for understanding their regulation (FIG. 3). Most promoters coincide with regions of high GC content and CpG ratios, or 'CpG islands'. These have been termed 'high CpG-content promoters' (HCPs), in contrast to 'low CpG-content promoters' (LCPs). Although HCPs and LCPs have different histone modification patterns and distinct modes of regulation[26,33], the distinction between HCPs and LCPs is somewhat arbitrary and does not effectively address several intermediate CpG content promoters. Incorporation of additional sequence features such as DNA motifs and DNA methylation patterns may result in a more precise and biologically meaningful classification[5,34]. Nonetheless, the two classes provide a useful framework for understanding and distinguishing the functions and regulation of mammalian promoters.

Initial ChIP followed by microarray (ChIP–chip) studies in mammalian cells revealed sharp peaks of histone H3 lysine 4 trimethylation (H3K4me3) associated with the TSSs of many transcribed genes[35,36] (FIG. 3). Subsequent studies of embryonic stem (ES) cell chromatin revealed surprisingly broad targeting of H3K4me3 to virtually all

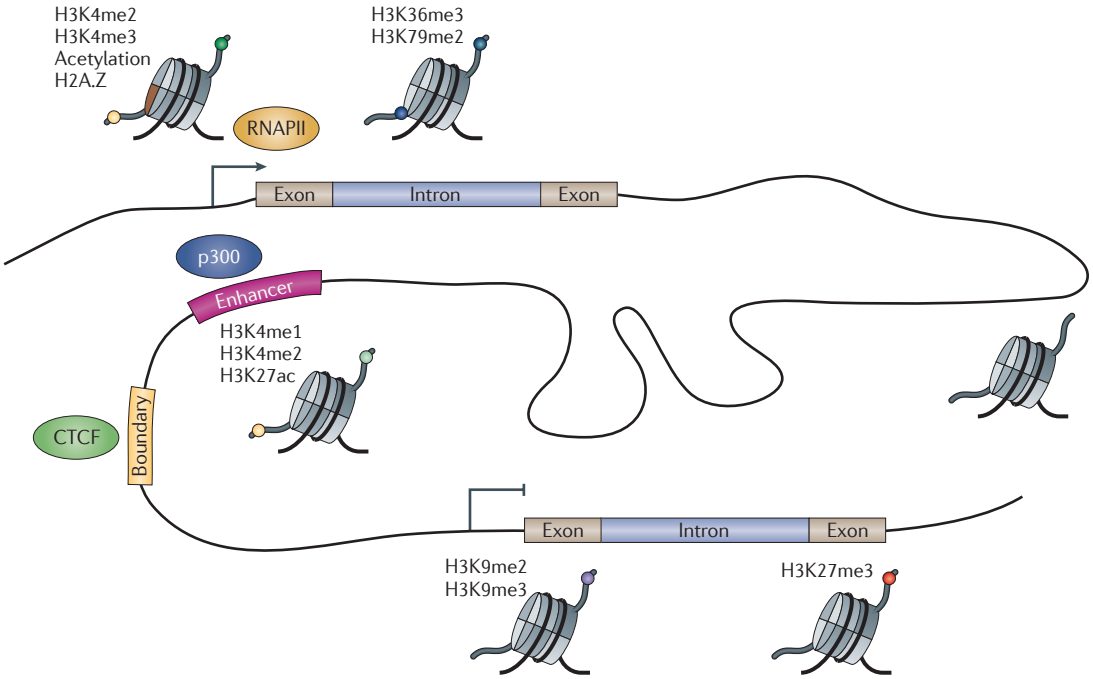

Figure 2 | **Histone modifications demarcate functional elements in mammalian genomes.** Promoters, gene bodies, an enhancer and a boundary element are indicated on a schematic genomic region. Active promoters are commonly marked by histone H3 lysine 4 dimethylation (H3K4me2), H3K4me3, acetylation (ac), and H2A.Z. Transcribed regions are enriched for H3K36me3 and H3K79me2. Repressed genes may be located in large domains of H3K9me2 and/or H3K9me3 or H3K27me3. Enhancers are relatively enriched for H3K4me1, H3K4me2, H3K27ac and the histone acetyltransferase p300. CTCF binds many sites that may function as boundary elements, insulators or structural scaffolds. These various features of chromatin help organize the DNA and distinguish functional elements in the large expanse of the genome. RNAPII, RNA polymerase II.

HCPs, regardless of expression state[24,26]. Sites of H3K4me3 were shown to be accompanied by other features of accessible chromatin, including histone acetylation, occupancy by the H3.3 histone variant and hypersensitivity to DNase I digestion[23,28,29,37]. Differentiated cells were also found to show relatively broad targeting of H3K4me3 to promoters, although with specific and biologically meaningful exceptions[26] (see below).

These accessible, H3K4me3-marked regions are also hypomethylated at the DNA level, as expected from their high CpG content[13,33]. This is consistent with a general exclusivity between such active and 'open' chromatin structures and DNA methylation. Indeed, several studies have provided evidence for direct antagonism between these epigenomic features. For instance, methylation of H3K4 was shown to preclude a physical interaction between the histone tail and DNA methyltransferase 3-like protein (DNMT3L)[38]. Another study, in the plant *Arabidopsis thaliana*, reported a direct role for H2A.Z — a histone variant enriched in genomic regions that are undergoing active nucleosome exchange — in protecting gene promoters from DNA methylation. In addition to a global exclusivity between sites of H2A.Z deposition and DNA methylation, this study also demonstrated that deficiency of H2A.Z deposition led to general DNA hypermethylation[39].

What mechanisms could underlie the correlation between these open chromatin features, H3K4me3 and the GC-rich promoters? ChIP–chip studies in ES cells showed that many H3K4me3-marked promoters are also enriched for RNA polymerase II (RNAPII) and subject to transcriptional initiation[24]. This was a surprising finding given that a substantial fraction of the HCPs does not produce detectable transcripts or undergo transcriptional elongation (see below). It suggests that transcriptional initiation and H3K4me3 are tightly linked and, moreover, that initiating RNAPII substantially contributes to the accessible chromatin configuration, potentially through interactions with chromatin modifiers as seen in yeast[7,40]. The concordance between H3K4me3 and HCPs may be more directly explained by the physical recognition of unmethylated CpG dinucleotides by CXXC domains in H3K4 methyltransferase complexes[41]. It was recently shown that introducing artificial, promoterless CpG clusters into mouse ES cells was sufficient to recruit the SET1 histone methyltransferase complex and establish H3K4me3 (REF. 42). A parallel study demonstrating targeting of an H3K36 demethylase complex by its CXXC domain suggests that such interactions may be general[43]. Together, these converging lines of experimental evidence suggest that transcriptional initiation and other pathways mutually reinforce a chromatin configuration that distinguishes this promoter class.

Regardless of the relative contributions of these proposed mechanistic models, the data suggest that HCPs tend to adopt an accessible chromatin state by default

Hidden Markov Model
A statistical model in which internal states are not visible but the outputs of these states are, and the outputs can therefore be used to infer the internal states. This model can be used to determine biologically relevant states from ChIP–seq data sets.

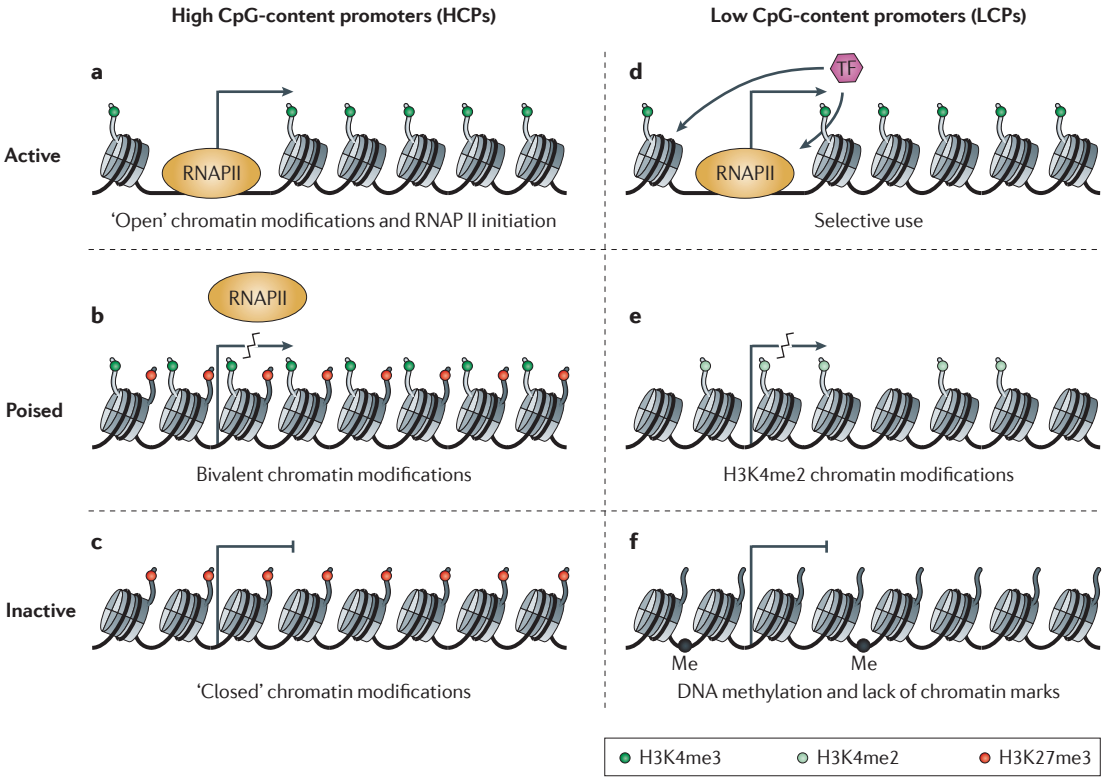

Figure 3 | **Chromatin patterns and regulation by promoter class.** Promoters can be classified according to their CpG content. High CpG-content promoters (HCPs) and low CpG-content promoters (LCPs) are subject to distinct chromatin patterns and regulation. **a** | HCPs have characteristics of accessible or 'active' chromatin by default. Active HCPs (for example, housekeeping gene promoters) are enriched for histone H3 lysine 4 trimethylation (H3K4me3) and subject to RNA polymerase II (RNAPII) initiation. They may be subject to additional regulation at the transition to elongation. **b** | Poised HCPs (for example, developmental regulator gene promoters in embryonic stem cells) are marked by the bivalent combination of H3K4me3 and H3K27me3. They may be subject to RNAPII initiation, but tend not to elongate or make productive mRNA. **c** | Inactive HCPs carry 'repressive' chromatin modifications such as H3K27me3 and are relatively inaccessible to RNAPII. Unlike HCP chromatin, LCP chromatin seems to be selectively activated (for example, by specific transcription factors (TFs)). **d** | Active LCPs are enriched for H3K4me3 and transcribed. **e** | Poised LCPs may be marked by H3K4me2 without H3K4me3. **f** | Inactive LCPs typically lack chromatin marks but may be DNA methylated (Me).

and are generally subject to a degree of transcription initiation. Thus, effective regulation of HCP genes is likely to require additional controls. Indeed, recent studies in macrophages and ES cells have documented roles for specific transcription factors in regulating steps downstream of initiation[44–46]. The research groups of Stephen Smale and Ruslan Medzhitov characterized a class of HCPs with constitutively active chromatin in macrophages that are basally transcribed by RNAPII, generating non-functional RNAs. After the macrophages are induced by lipopolysaccharide, the transcription factor nuclear factor-κB (NF-κB) initiates a cascade that causes RNAPII to adopt a more processive form (that is, its carboxy-terminal domain becomes phosphorylated at serine 2) and results in the rapid production of functional transcripts[44,45]. In ES cells, genome-wide-mapping studies revealed a key role for the transcription factor MYC in enhancing the 'release' of RNAPII at HCPs and, hence, promoting the generation of mature transcripts[46]. Together, these studies emphasize the importance and complexity of downstream steps in controlling the expression of genes associated with this major promoter class.

In marked contrast to HCPs, LCPs seem inactive by default (FIG. 3). Indeed, most annotated LCPs lack H3K4me3 (or H3K4me2) in ES cells and in various differentiated cell types[26,33]. The minority of LCPs that are marked by H3K4me3 seem to be fully expressed with the levels of transcripts from these promoters being substantially higher than their unmarked counterparts.

Further biological insights into LCP regulation emerged from an analysis of chromatin structure changes during haematopoietic differentiation[47]. Orford *et al.* defined a subset of promoters that carry H3K4me2 but not H3K4me3 in haematopoietic progenitors. They found that this set corresponded to LCPs associated with haematopoietic cell type-specific genes that are generally inactive in progenitors but become induced during differentiation. Specifically, they observed a switch from H3K4me2 to H3K4me3 on induction of such LCPs during differentiation. These studies suggest that LCPs are subject to greater regulation at the level of transcription initiation, and may be poised in certain contexts by lower degrees of histone methylation. Notably, genes subject to this form of regulation tend to encode proteins

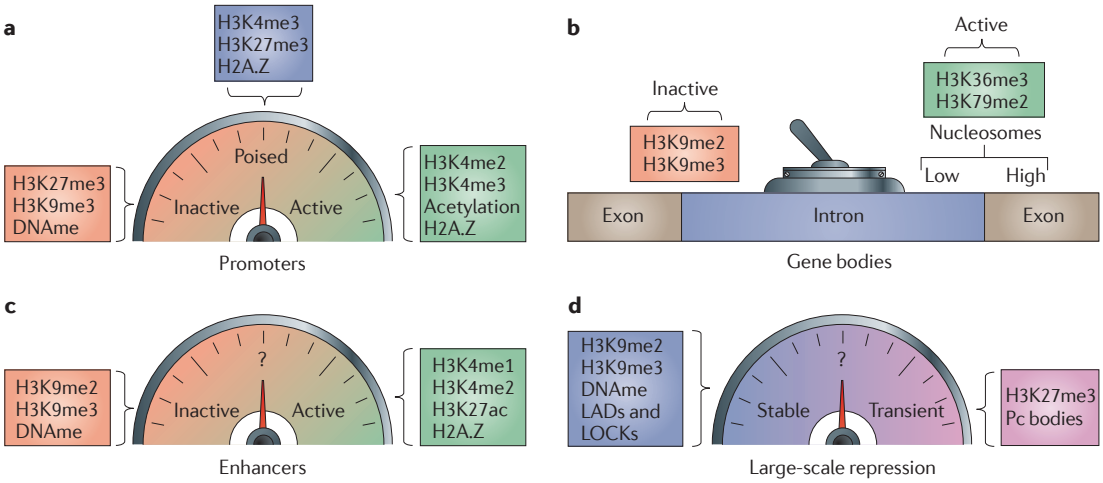

Figure 4 | **'Dashboard' of histone modifications for fine-tuning genomic elements.** In addition to enabling annotation, histone modifications may serve as 'dials' or 'switches' for cell type specificity. **a** | At promoters, they can contribute to fine-tuning of expression levels — from active to poised to inactive — and perhaps even intermediate levels. **b** | At gene bodies, they discriminate between active and inactive conformations. In addition, exons in active genes have higher nucleosome occupancy and thus more histone H3 lysine 36 trimethylation (H3K36me3) and H3K79me2-modified histones than introns. **c** | At distal sites, histone marks correlate with levels of enhancer activity. **d** | On a global scale, they may confer repression of varying stabilities and be associated with different genomic features. For example, lamina-associated domains (LADs) in the case of stable repression and Polycomb (Pc) bodies in the case of context-specific repression. DNAme, DNA methylation; LOCK, large organized chromatin K modification.

specific to terminally differentiated cells (for example, structural proteins) instead of the regulators that drive cell fate (for example, developmental transcription factors). The regulatory genes involved in determining cell fate have HCPs and are subject to more complex regulation by Polycomb complexes (see below).

Poised and repressed chromatin states. Repressed promoters also show unique patterns of chromatin modifications that seem to reflect distinct modes of transcriptional silencing. These include H3K27me3, the prototypical mark of Polycomb repressors; H3K9me3, which correlates with constitutive heterochromatin; and DNA methylation (FIG. 4a).

Polycomb proteins are transcriptional repressors essential for maintaining tissue-specific gene expression programmes in multicellular organisms[6]. In mammals, a large proportion of HCPs is targeted by the main Polycomb repressive complexes — Polycomb repressive complex 1 (PRC1) and PRC2. In ES cells, approximately 20% of HCPs are bound by PRC2 and marked by the associated modification, H3K27me3 (REFS 26,48–52). These promoters have been termed 'bivalent' as they also carry H3K4me3 and thus have characteristics of both activating and repressive chromatin[53,54]. Bivalent, PRC2 target promoters have attracted considerable interest, as a large proportion corresponds to developmental genes that encode transcription factors and other regulators of cellular state. These genes are largely inactive in pluripotent cells, but can be rapidly induced or stably inactivated, depending on the developmental course. It has been proposed that the signature chromatin configuration is instrumental for poising bivalent promoters for these alternate fates. Indeed, global studies of neural and haematopoietic progenitors indicate that bivalent chromatin tends to resolve at successive developmental stages in a pattern that closely matches the expression state and future potential of the corresponding genes[26]. For example, Mohn *et al.* followed H3K27me3 patterns in gene promoters during the transition of ES cells to neural progenitors and subsequently to terminally differentiated neurons, and found a progression of HCP modifications in accordance with expression state and gene potential[55]. Similar patterns are also evident along the axis of haematopoietic differentiation, as indicated by the analysis of *in vivo* lineages from both human and mouse[56,57].

Although bivalent promoters in ES cells have very low expression levels and were initially found to be free from RNAPII[55], subsequent studies have suggested that at least a subset has detectable RNAPII enrichment[24,58]. This raises the possibility that initiating RNAPII contributes to the establishment of H3K4me3, or potentially even H3K27me3, at these loci. However, these data should be interpreted with some caution. RNAPII enrichment was only detected under certain experimental conditions[58] and, moreover, evidence for RNA transcription at these loci remains scarce[59]. Other technical issues of possible relevance include an inherent promoter bias in some ChIP data and possible heterogeneity of the cell population studied owing to partial differentiation.

How is PRC2 targeted to HCPs? The GC-rich sequences of HCPs are likely to play an important part, given the strong correlation between CpG islands and PRC2 binding. PRC2 targets in ES cells can be predicted with remarkable accuracy by simply identifying CpG islands depleted of motifs for activating transcription factors[48]. A causal role for such CpG sequences is

supported by the finding that introduction of exogenous GC-rich sequence elements into ES cells is sufficient to mediate PRC2 recruitment[60]. Still, the underlying mechanisms are not yet understood. Although sequence-specific DNA-binding proteins guide PRC2 to target elements in *D. melanogaster*, analogous factors have yet to be identified in mammals. Rather, mammalian PRC2 contains the atypical DNA-binding proteins adipocyte enhancer-binding protein 2 (AEBP2)[61] and JARID2 (REFS 62–65). JARID2 was recently shown to be essential for PRC2 function and the establishment of proper H3K27me3 patterns[62–65]. ChIP–seq analysis confirmed that JARID2 colocalizes with PRC2 and H3K27me3 at GC-rich sequence elements. However, *in vitro* biochemical studies suggest that JARID2 is a promiscuous DNA-binding protein without particular specificity for GC-rich sequences[66]. Hence, it does not seem that this factor can fully explain PRC2 targeting. Non-coding RNAs have also emerged as intriguing candidates for PRC2 recruitment. PRC2 has an affinity for various RNA classes such as short GC-rich RNAs that might have a role in targeting the complex to weakly initiating HCPs[67]. PRC2 can also interact with long intergenic non-coding RNAs (lincRNAs), including *Xist* and *HOTAIR*, both of which seem to play important parts in the localization and stabilization of Polycomb complexes in differentiating cells[68,69]. PRC2 association is further stabilized by its own affinity for K27-methylated H3 tails[70,71]. Thus, in contrast to *D. melanogaster*, PRC2 localization in mammals seems to be directed to GC-rich elements by a complex interplay between low specificity DNA-binding proteins, RNA-targeting factors and chromatin-based stabilization.

The challenge of understanding Polycomb localization is further complicated by PRC1, a repressive complex that ubiquitylates histone H2A and may also mediate the structural compaction of chromatin[6]. In ES cells, PRC1 associates with a specific subset of PRC2 targets that includes key developmental regulators and other genes subject to epigenetic repression during development[48]. These PRC1 targets tend to have larger CpG islands or extended GC-rich regions relative to PRC2-specific loci. In addition, recent studies have identified specific DNA elements that contain binding motifs for the transcriptional repressor protein YY1 that can initiate PRC1-dependent silencing during development[72,73]. A unifying theory for how Polycomb complexes are targeted is an important goal, as both PRC1 and PRC2 are almost certainly required for stable epigenetic gene repression[6].

The landscape of Polycomb repression changes markedly through differentiation. In addition to the progressive resolution of bivalent chromatin at specific promoters described above, a smaller subset of promoters is subject to *de novo* gain of H3K27me3 during development[55]. The affected genes include certain pluripotency regulators repressed during ES cell differentiation[51]. At many loci, differentiation is also accompanied by dramatic spreading of the histone modification to yield contiguous but more diffuse domains of H3K27me3 (REF. 74).

Relatively less is known about the role of DNA methylation in HCP regulation during development. Hypermethylation of individual CpG islands and extended genomic loci have been widely described in human cancer[75,76], yet genome-scale studies suggest that most CpG islands remain largely unmethylated during normal development[13,55]. However, closer inspection of the DNA methylation pattern of HCPs shows that although the CpG islands are unmethylated, their 'shores' — sequences up to 2 kb distant from the CpG islands — frequently become methylated in tissue-specific patterns[77]. CpG island shores may also be conserved between human and mouse and, when methylated, correlate with gene silencing in a tissue-specific manner. Although the functionality of CpG shores remains controversial, global reduction of DNA methylation by a small molecule (5-azacytidine) or by knockout of DNA methyltransferases shows concurrent activation of these genes. More broadly, genome-scale and genome-wide analyses of DNA methylation patterns have provided insights into ES cell regulation[14], haematopoietic differentiation[78] and epigenetic roadblocks to cellular reprogramming[79] (for in-depth reviews, see REFS 3, 77, 78).

Up to 80% of LCPs are DNA methylated in ES cells[13,80]. The functional consequence of this DNA methylation remains unclear; the relative paucity of CpG dinculeotides in these regions suggests that the impact of methylation may be slight. Interestingly, however, inactive LCPs are frequently located in extended regions of H3K27me3 or H3K9me3 that may reflect large-scale sequestration of inactive genomic regions, and thereby hold the potential for contextual repression of chromosomal regions (see below).

Gene bodies. Mammalian genes are characterized by large numbers of exons in an expanse of introns. In many cases, alternative splicing provides an additional layer of complexity and regulation[81]. Recent studies suggest that, at the DNA level, chromatin patterns can distinguish primary transcripts and exons, and may even have a role in determining splicing patterns. Major marks seen in transcribed regions include H3K36me3 (REFS 22,26) and H3K79me2 (REF. 7) (FIG. 4b). Comparative analyses of H3K36me3 with expression and splicing data reveal several emerging trends. First, H3K36me3 levels correlate with levels of gene expression[22,26]. This is likely to reflect interactions between elongating RNAPII and the methyltransferases that deposit this mark[7].

Recent studies have noted that expressed exons have particularly strong enrichment for H3K36me3 (REFS 82–84) compared with introns. They may also show modest enrichment for H2BK5me1, H4K20me1 and H3K79me1 (REF. 29). Subsequent studies have indicated that the observed enrichments for histone marks likely reflect the preferential occupancy and positioning of nucleosomes over exons[82,85] (FIG. 4b). Specifically, computational analyses in these studies suggest that this higher abundance of nucleosomes might account for the observed exonic H3K36me3 enrichment. The authors of

these studies speculated that positioned nucleosomes at exons might enhance splicing by acting as 'speed bumps' to slow RNAPII. According to this model, the splicing machinery is recruited during transcription and an increased RNAPII occupancy time might translate into improved recognition of splicing signals[86].

A recent study by Tom Misteli's group more directly linked histone modifications at gene bodies with the splicing machinery[87]. These authors studied the alternatively spliced gene fibroblast growth factor receptor 2 (*FGFR2*) and found that histone modifications across the gene vary among cell types. Specifically, they observed distinct patterns of H3K36me3, H3K4me3, H3K4me1 and H3K27me3 across *FGFR2* in epithelial cells and mesenchymal cells, which produce different splice forms. Remarkably, by modulating the levels of H3K36me3 and H3K4me3, the authors could influence the splicing patterns of *FGFR2*. They suggest a model in which histone marks are read by the splicing machinery through the histone tail-binding protein mortality factor 4-like protein 1 (MORF4L1) and the splicing regulator polypyrimidine tract-binding protein 1 (PTBP1). Interestingly, if these histone patterns are general signatures of alternatively spliced exons, a comparison of genome-wide maps of these marks in different cell types might reveal global maps of alternative splicing events. Regardless, the robust enrichment of modified nucleosomes at exons suggests that a link between histone modifications and splicing may be a general phenomenon.

Enhancers. Enhancers are DNA elements that recruit transcription factors, RNAPII and chromatin regulators to positively influence transcription at distal promoters[88]. Histone modification profiles have proven to be particularly useful for identifying enhancer elements in an unbiased fashion. In addition to specific histone modifications, enhancers are preferentially occupied by sequence-specific DNA-binding proteins[27] and co-activators such as p300 (REF. 89). By observing the histone modifications at distal p300-binding sites, Heintzman *et al.* identified relative H3K4me1 enrichment and H3K4me3 depletion as a chromatin signature of enhancers in human cells[25]. The group used this signature to predict over 55,000 candidate enhancers in five human cell types, including K562 and HeLa cells[90]. Interestingly, the chromatin patterns at enhancers were much more variable and cell type specific than chromatin patterns at promoters or insulators. This study suggested a crucial role for enhancers in controlling the level and timing of gene expression in a cell type-specific manner and highlighted the power of histone modification profiling for identifying diverse functional elements.

Despite the fruitful application of a histone modification signature to predict enhancers, the mechanism by which H3K4me1 is established at these sites remains unknown. Integrative analyses suggest that enhancers also share enrichment for H3K27 acetylation, H2BK5me1, H3K4me2, H3K9me1, H3K27me1 and H3K36me1, suggesting redundancy in the histone marks[28]. This signature might indicate general genome accessibility or chromatin dynamics at these sites. It might also reflect the physical proximity of enhancer elements to activating chromatin machinery at their target promoters through looping interactions[88]. The chromatin patterns at enhancers could also be actively fine-tuned, as different patterns of acetylation and H2A.Z deposition correlate with differences in downstream gene expression levels[29] (FIG. 4c).

Support for a more direct interaction between enhancers and the transcriptional machinery emerged from a recent genome-wide study that mapped p300 and H3K4me1 in mouse cortical neurons. Kim *et al.* found that RNAPII interacts with many active enhancers that were identified by the chromatin patterns in these cells and transcribes bidirectional short (<2 kb) non-coding RNAs, termed enhancer RNAs (eRNAs)[91]. The expression levels of eRNAs correlate with the proximal gene activity, and eRNA synthesis seemed to require interaction with the relevant promoter. The function of these eRNAs is not understood, but similar findings also emerged from a study of enhancer elements in macrophages[92]. Transcription of eRNAs might be needed to maintain open chromatin at the enhancer region but, alternatively, might be a byproduct of the chromatin configuration or looping.

Insulators and boundary elements. Insulators are DNA elements that block enhancer activities[93] (FIG. 2). They are likely related to boundary elements, which are defined by their capacity to prevent heterochromatin spreading. In mammals, the transcriptional repressor CTCF has been implicated in blocking of enhancer activity and heterochromatin spreading, and in interchromosomal and intrachromosomal organization. CTCF has been profiled genome-wide in several human cell types, revealing tens of thousands of binding sites in primary human fibroblasts, $CD4^+$ T cells and HeLa cells[22,90,94]. These studies have come to the consensus that most CTCF-binding sites share a common motif and are relatively invariant across different cell types. The CTCF-binding sites also show modest enrichment for the histone variant H2A.Z but, surprisingly, vary widely in terms of other histone modifications[28,29]. Recent models suggest that CTCF, most likely in association with cohesin[95], can stabilize long-range DNA interactions and chromatin loops. In this way, the factor is thought to be instrumental in establishing a defined three-dimensional genome structure and partitioning distinct chromatin domains[93].

Higher-order chromatin organization

As cells differentiate from a totipotent to a specialized committed state, a high percentage of their genome must be stably repressed. In this regard, chromatin regulators and histone modifications seem to work in conjunction with other mechanisms to silence broad genomic regions. There are several known modes of large-scale repression that correlate with megabase (Mb) domains of H3K9me3 and H3K27me3 (FIG. 4d) and likely reflect specialized higher-order chromatin structures in the nucleus (FIG. 5).

H3K9me2 and lamina-associated domains. The nuclear lamina is thought to bind and silence large regions of heterochromatin. Two studies that analysed distinct genomic features identified similar sets of domains enriched for H3K9 methylation and lamina contact[96,97]. Guelen *et al.* globally mapped the interaction between the genome and nuclear lamina in human fibroblasts using DamID. These authors observed two discrete chromatin environments: lamina-associated domains (LADs) and regions outside LADs. Both regions were approximately 0.1–10 Mb in size. LADs were found to have low gene density, low transcriptional activity and a paucity of active chromatin modifications. Although the nuclear lamina had previously been associated with inactivity, for the first time, these studies defined the locations and extents of LADs and the correlated chromatin patterns. Remarkably, tethering experiments show that interaction with the nuclear lamina is not only correlative but is also causal in reducing gene expression[98–100].

Wen *et al.* identified a similar set of genomic domains by analysing genome-wide maps of H3K9me2 in differentiated and undifferentiated cells[97]. They found large and diffuse regions of K9 methylation that cover up to 4.9 Mb and collectively represent up to 46% of the genome, which they termed large organized chromatin K modifications (LOCKs). These investigators also showed that LOCKs are conserved between human and mouse, and that the H3K9me2 mark was dependent on the G9A H3K9 methyltransferase. Furthermore, a close relationship between LOCKs and LADs was indicated by a striking overlap of 82% between placental LOCKs and LADs found in fibroblasts. Thus, genomic regions diffusely marked by H3K9 methylation seem to be in contact with the nuclear lamina; these findings have prompted a model in which chromatin is partitioned into distinct environments in different cell types. It was initially proposed that LOCKs are relatively scarce in ES cells, as few such chromatin domains could be detected. However, whether this reflects a true distinction in modification patterns between cell types or a detection bias has been questioned[101]. The nature of these compartments remains an area of active investigation, as these structures could play a crucial part in sequestering unused regions of the genome, and thereby reducing the effective 'search space' for gene regulatory machinery.

H3K27me3 blocks and Polycomb bodies. Genome-wide histone modification maps have also revealed large blocks of H3K27me3 in differentiated cells. Identification of these domains relied on new algorithms for identifying broad regions — rather than sharp peaks — of enrichment, as two recent studies illustrate. Pauler *et al.* used an algorithm called broad local enrichments (BLOCs) to identify H3K27me3 blocks that are on average 43 kb and overlap silent genes and intergenic regions[102]. They found this pattern in numerous ChIP–chip and ChIP–seq data sets, and suggest that this is a common feature of H3K27me3 in differentiated cell types. The authors speculate that these H3K27me3 blocks may relate to Giemsa bands, as they observe alternating chromatin patterns along chromosomes. Hawkins *et al.* used ChromaBlocks to find similar H3K27me3 blocks in human IMR90 fibroblasts and characterized their dynamics during differentiation[74]. This study suggested that these repressive domains are often seeded in ES cells and expand in differentiated cell types, apparently to confer cell type-specific repression (FIG. 4d). As these domains have only recently been observed, little is known about their establishment or functional consequences. It is tempting to consider the possibility that, like H3K9me2 domains, H3K27me3 blocks mark distinct nuclear structures or regions. They potentially correspond to Polycomb bodies, which are discrete foci of silenced genes that have been observed by imaging and *in situ* hybridization in fly and human cells[103]. Although there are no data yet that directly link H3K27me3 blocks to these structures, there is indirect evidence of H3K27me association with compacted chromatin; H3K27me3 can promote recruitment of PRC1 (REF. 6), and PRC1 may be required

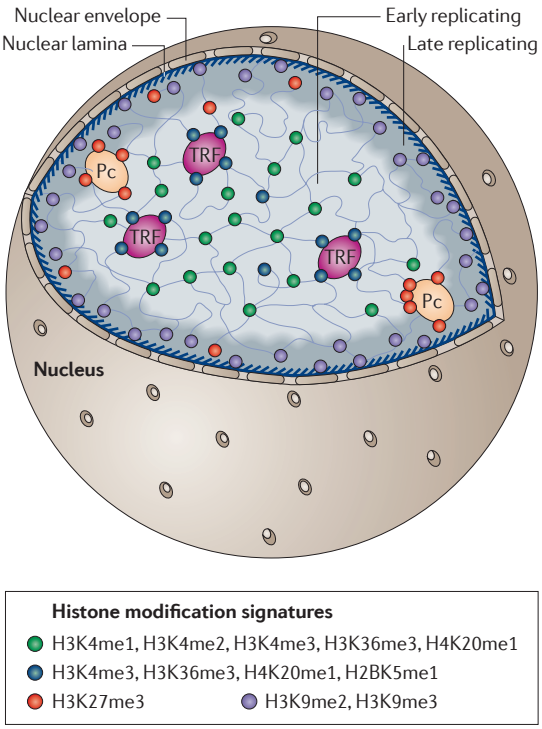

Figure 5 | **Histone modification signatures associated with features in the mammalian cell nucleus.** Signature histone modifications correlate with various nuclear features, although the relationships might be indirect. Chromatin with modifications generally associated with active transcription (green dots) often replicates early, whereas chromatin with generally repressive modifications (purple dots) replicates late. Regions enriched for some sets of active modifications (blue dots) may converge into transcription factories (TRFs). Blocks of histone H3 lysine 27 trimethylation (H3K27me3; red dots) may form Polycomb bodies (Pc) and diffuse domains marked by H3K9me2 or H3K9me3 (purple dots) may contact the nuclear lamina.

DamID
A method for mapping the distribution of chromatin-associated proteins by fusing a protein of interest with *E. coli* DNA adenine methyltransferase (Dam), which methylates adenines proximal to the binding sites of a protein, thus circumventing the need for antibodies.

Giemsa band
Also known as a G-band. A characteristic banding pattern is obtained by treating chromosomes with Giemsa stain. The intensity of Giemsa staining is correlated with genomic features. For instance, dark Giemsa bands usually are AT rich, have low gene density and have higher densities of repeat elements.

Polycomb body
A discrete nuclear focus containing Polycomb proteins and their silenced target genes. Polycomb bodies have been observed in *D. melanogaster* and human cells by imaging and *in situ* hybridization.

to maintain chromatin compaction at the Hox loci in ES cells[104]. Together, these studies support connections between Polycomb regulation, histone modifications and chromatin compartmentalization that promise to be an exciting area for further investigation.

Replication time zones. In addition to delineating particular genomic elements, chromatin patterns gleaned through mapping studies also seem to relate to DNA replication timing (FIG. 5). The genome has distinct replication time zones that are on average 1 Mb in size and tend to undergo DNA synthesis at coordinated times during S phase[105]. Plasmid injection experiments initially suggested a tight link between replication timing and histone H3 and H4 acetylation. Regardless of sequence, a DNA fragment that is introduced into a cell in early S phase will be wrapped around acetylated histones, however, the same fragment will be associated with deacetylated histones when injected in late S phase[106]. Genome-wide profiling of replication timing in mouse and human cells revealed a correlation between replication domains and chromatin structure[107,108]. Early replicating zones associate with H3K4me1, H3K4me2, H3K4me3, H3K36me3, H4K20me1, and H3K9 and H3K27 acetylation, whereas late replicating zones mostly correlate with H3K9me2, and to a lesser degree with H3K9me3 (REF. 107). Subsequent studies have shown that the relationship between replication domains and histone modifications can be more than correlative, as histone acetylation patterns directly influence the time at which origins initiate replication ('fire') during S phase in yeast and mouse models[109,110]. Of note, bivalent chromatin replicates early despite being transcriptionally inactive, potentially reflecting its accessible and poised character[53]. Also, boundaries between replicating zones have a signature modification pattern — peaks of H3K4me1, H3K4me2, HeK4me3, H3K27ac and H3K36me3. It has been speculated that the 'active' histone modifications might serve as boundary elements that block spreading of late-replicating heterochromatin. Together, these studies above illustrate global links between histone modification patterns, replication timing and higher-order nuclear structures (FIG. 5).

Perspectives and future challenges

The growing panel of genome-wide histone modification maps has several implications. At the level of the primary chromatin structure, the data suggest that histone modifications indicate functional genomic elements, gene expression, splicing patterns and modes of repression. Together with studies that perturb the mechanisms that write and read these marks, this insight may enable us to better understand and predict how normal or diseased cell types use and regulate their genomes. Additionally, these maps promote an appreciation of the three-dimensional organization of the genome. During the past few years, more pieces of the nuclear architecture 'jigsaw puzzle' have been revealed. As we have discussed, histone modifications are intimately tied to large-scale repressive domains like LADs and Polycomb bodies, and broad patterns of replication time zones. Together with ongoing studies of additional structures, such as transcription factories and nucleolus-associated domains[111,112], these findings are building a better understanding of the architecture of chromatin in the nucleus.

Several recent technological advances direct us towards a molecular understanding of chromatin spatial organization. Lieberman-Aiden *et al.* scaled the chromosome conformation capture (3C) assay for unbiased genome-wide identification of chromatin interactions (Hi-C)[113]. This approach revealed distinct spatial compartments distinguished by their degree of openness, but was limited in terms of the resolution with which it could distinguish interactions and compartmentalization. Fullwood *et al.* scaled the technology of a related approach that also incorporates an immunoprecipitation step (chromatin interaction analysis using paired-end tag sequencing; ChIA–PET)[114]. They focused on the interaction network bound by oestrogen receptor-α, and noted numerous cases of chromatin looping for coordinated transcriptional regulation. Another important area of technology development relates to miniaturization and increasing the sensitivity of assays so they may be compatible with small samples or even individual cells[56,115]. High-resolution-imaging approaches may also be instrumental in this regard. Combined with more powerful and integrative computational algorithms, such tools should ultimately enable every genomic region in a living cell to be tracked across differentiation, development and disease.

Despite our increasing knowledge on various aspects of chromatin structure, we are still far from understanding the determinants of this structure. Relatively little is known about the complexes that introduce and maintain histone modification patterns. Even less is known about the way specific modification signatures or 'states,' are read. How combinatorial options of chromatin 'writer' and 'reader' proteins facilitate more sophisticated and robust regulation of gene expression and genome function remains a key area of investigation. Detailed knowledge of global chromatin architecture, along with these regulators, represents a crucial step towards understanding how genetic, epigenetic, and environmental or stochastic factors drive context-specific genome regulation.

This era is an exciting time in biology, in which new genomic tools are validating or refuting dogmas developed through gene-specific analysis, as well as illuminating entirely unexpected principles. The pace of change is accelerating thanks to remarkable advances in DNA sequencing, the increasing availability of epigenomic data in the public domain from the National Institutes of Health and international projects[27,116,117], and the rapid dissemination of these technologies into individual research laboratories. By changing our focus from 'gene-centred' to 'genome-wide', such approaches hold much promise to enhance our understanding of genome architecture and its consequences on gene regulation, genome stability, cell phenotype and organismal physiology in both health and disease.

3C
Chromosome conformation capture is a method to map chromosome interactions locally. It relies on an increased frequency of intramolecular ligation between fragments in close three-dimensional proximity in the nucleus.

1. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
2. Schones, D. E. & Zhao, K. Genome-wide approaches to studying chromatin modifications. *Nature Rev. Genet.* **9**, 179–191 (2008).
3. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11**, 204–220 (2010).
4. Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nature Rev. Genet.* **11**, 285–296 (2010).
5. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
6. Simon, J. A. & Kingston, R. E. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nature Rev. Mol. Cell Biol.* **10**, 697–708 (2009).
7. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
8. Jirtle, R. L. & Skinner, M. K. Environmental epigenomics and disease susceptibility. *Nature Rev. Genet.* **8**, 253–262 (2007).
9. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
10. Hesselberth, J. R. *et al.* Global mapping of protein–DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6**, 283–289 (2009).
11. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
12. Auerbach, R. K. *et al.* Mapping accessible chromatin regions using Sono–Seq. *Proc. Natl Acad. Sci. USA* **106**, 14926–14931 (2009).
13. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
14. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
15. Down, T. A. *et al.* A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotech.* **26**, 779–785 (2008).
16. Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
17. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
18. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
19. Hawkins, R. D., Hon, G. C. & Ren, B. Next-generation genomics: an integrative approach. *Nature Rev. Genet.* **11**, 476–486 (2010).
20. Park, P. J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Rev. Genet.* **10**, 669–680 (2009).
21. Solomon, M. J., Larsen, P. L. & Varshavsky, A. Mapping protein–DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937–947 (1988).
22. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
This pioneering study highlighted the value of comprehensive and high-throughput sequencing approaches to map histone modifications. The data generated have been extensively analysed by many other groups and used to generate hypotheses and models on chromatin function.
23. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40**, 897–903 (2008).
24. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
25. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
26. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
This was among the first studies to apply high-throughput sequencing to map chromatin. Maps for ES and differentiated cells provided broad views of the chromatin changes that accompany cellular commitment.
27. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
28. Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **5**, e1000566 (2009).
29. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotech.* **28**, 817–825 (2010).
References 28 and 29 present innovative approaches for integrating genome-wide chromatin data sets. The algorithms described result in systematic insights into the roles of and interrelationships among histone modifications, and provide a framework for handling the increasing volumes of epigenomic data now being produced.
30. Dion, M. F. *et al.* Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl Acad. Sci. USA* **102**, 5501–5506 (2005).
31. Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. Yeast histone H4 N-terminal sequence is required for promoter activation *in vivo*. *Cell* **65**, 1023–1031 (1991).
32. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
33. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
34. Straussman, R. *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nature Struct. Mol. Biol.* **16**, 564–571 (2009).
35. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
36. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
37. Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678–691 (2010).
38. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**, 714–717 (2007).
39. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).
This study suggests a direct role for H2A.Z. in protecting gene promoters from DNA methylation. In addition to the general exclusivity between sites of H2A.Z deposition and DNA methylation, it could be demonstrated that H2A.Z deficiency leads to broad DNA hypermethylation.
40. Shilatifard, A. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.* **20**, 341–348 (2008).
41. Lee, J. H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3–Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* **280**, 41725–41731 (2005).
42. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
43. Blackledge, N. P. *et al.* CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* **38**, 179–190 (2010).
44. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).
45. Hargreaves, D. C., Horng, T. & Medzhitov, R. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**, 129–145 (2009).
46. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
47. Orford, K. *et al.* Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Dev. Cell* **14**, 798–809 (2008).
48. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242 (2008).
49. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
50. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
51. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**, 299–312 (2007).
52. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298 (2007).
53. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532–538 (2006).
54. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
55. Mohn, F. *et al.* Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755–766 (2008).
56. Adli, M., Zhu, J. & Bernstein, B. E. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nature Methods* **7**, 615–618 (2010).
57. Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80–93 (2009).
58. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biol.* **9**, 1428–1435 (2007).
59. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
60. Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* (in the press).
61. Kim, H., Kang, K. & Kim, J. AEBP2 as a potential targeting protein for Polycomb repression complex PRC2. *Nucleic Acids Res.* **37**, 2940–2950 (2009).
62. Li, G. *et al.* Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* **24**, 368–380 (2010).
63. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310 (2010).
64. Peng, J. C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290–1302 (2009).
65. Shen, X. *et al.* Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**, 1303–1314 (2009).
66. Kim, T. G., Kraus, J. C., Chen, J. & Lee, Y. JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. *J. Biol. Chem.* **278**, 42247–42255 (2003).
67. Kanhere, A. *et al.* Short RNAs are transcribed from repressed Polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
68. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
69. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
70. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
71. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biol.* **10**, 1291–1300 (2008).
72. Sing, A. *et al.* A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* **138**, 885–897 (2009).
73. Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J. & Kingston, R. E. A region of the human HOXD cluster that confers Polycomb-group responsiveness. *Cell* **140**, 99–110 (2010).

74. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
75. Coolen, M. W. *et al.* Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nature Cell Biol.* **12**, 235–246 (2010).
76. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
77. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41**, 178–186 (2009).
78. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338–342 (2010).
79. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
80. Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).
81. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
82. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
83. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).
84. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
85. Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).
References 82 and 85 describe computational analysis of published ChIP–seq data, and present evidence for higher nucleosome abundance at exons compared to introns.
86. Kornblihtt, A. R., Schor, I. E., Allo, M. & Blencowe, B. J. When chromatin meets splicing. *Nature Struct. Mol. Biol.* **16**, 902–903 (2009).
87. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
This study was the first to directly link histone modifications at gene bodies with the splicing machinery. The authors show that distinct patterns of histone modifications across an alternatively spliced gene vary between cell types along with changes in its splice forms.
88. Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
89. Visel, A. *et al.* ChIP–seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
90. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
Building on previous work that introduced the use of chromatin signatures to predict enhancers, this group showed that chromatin patterns at enhancers are more cell type specific than those at promoters.
91. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
92. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
93. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
94. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
95. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
96. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
This was one of the first papers to provide a global view of higher-level genome organization by mapping megabase-scale regions associated with lamina.
97. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genet.* **41**, 246–250 (2009).
This paper provided evidence that large domains of H3K9me2 organize inactive chromatin and are altered in differentiation.
98. Finlan, L. E. *et al.* Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **4**, e1000039 (2008).
99. Kumaran, R. I. & Spector, D. L. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J. Cell Biol.* **180**, 51–65 (2008).
100. Reddy, K. L., Zullo, J. M., Bertolino, E. & Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243–247 (2008).
101. Filion, G. J. & van Steensel, B. Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. *Nature Genet.* **42**, 4 (2010).
102. Pauler, F. M. *et al.* H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* **19**, 221–233 (2009).
103. Sexton, T., Schober, H., Fraser, P. & Gasser, S. M. Gene regulation through nuclear organization. *Nature Struct. Mol. Biol.* **14**, 1049–1055 (2007).
104. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–64 (2010).
105. Goren, A. & Cedar, H. Replicating by the clock. *Nature Rev. Mol. Cell Biol.* **4**, 25–32 (2003).
106. Zhang, J., Xu, F., Hashimshony, T., Keshet, I. & Cedar, H. Establishment of transcriptional competence in early and late S phase. *Nature* **420**, 198–202 (2002).
107. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
108. Karnani, N., Taylor, C., Malhotra, A. & Dutta, A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome Res.* **17**, 865–876 (2007).
109. Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B. J. & Grunstein, M. Histone acetylation regulates the time of replication origin firing. *Mol. Cell* **10**, 1223–1233 (2002).
110. Goren, A., Tabib, A., Hecht, M. & Cedar, H. DNA replication timing of the human β-globin domain is controlled by histone modification at the origin. *Genes Dev.* **22**, 1319–1324 (2008).
111. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42**, 53–61 (2010).
112. Nemeth, A. *et al.* Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).
113. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
This paper introduced a new technology for unbiased detection of genome interactions. The authors used the data generated to reconstruct the three-dimensional structure and organization of the genome.
114. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
This paper introduced a new technology for the unbiased genome-wide detection of chromatin interactions and focused on the regulatory targets of oestrogen receptor-α.
115. Goren, A. *et al.* Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. *Nature Methods* **7**, 47–49 (2010).
116. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotech.* **28**, 1045–1048 (2010).
117. Satterlee, J. S., Schubeler, D. & Ng, H. H. Tackling the epigenome: challenges and opportunities for collaboration. *Nature Biotech.* **28**, 1039–1044 (2010).

Acknowledgements
We thank E. Mendenhall, M. Ku, R. Koche and E. Rheinbay for critical reading of the manuscript. We also thank members of the Bernstein laboratory for insightful discussions. V.W.Z. was supported by a National Defense Science and Engineering Graduate Fellowship and a National Science Foundation Graduate Research Fellowship. A.G. was supported by an EMBO long-term postdoctoral fellowship. B.E.B. is an Early Career Scientist of the Howard Hughes Medical Institute. Research in the Bernstein laboratory is supported by funds from the Burroughs Wellcome Fund, Howard Hughes Medical Institute and the National Institutes of Health.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Nature Reviews Genetics article series on Applications of next-generation sequencing:
http://www.nature.com/nrg/series/nextgeneration/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF